Dr. Ron Brachman, Director
Information Processing Technology Office (IPTO)
Systems that Know What They're Doing: The New DARPA/IPTO Initiative in Cognitive Systems

Good morning. I am very pleased to be here with you this morning and I am especially excited to have the opportunity to share with you some thoughts about a dramatic new direction for our work in computing and information processing.

As the video just shown makes clear, the DARPA office created in the early '60s by Jack Ruina and first headed by J.C.R. Licklider—originally called the Information Processing Techniques Office, or IPTO—was responsible for some of the most dramatic changes of the 20th century in American technology.

The IPTO legacy is familiar: time-sharing, interactive computing, the essential thinking that made the computer personal, the ARPANet, ILLIAC 4, the Internet, and many other examples. Both in national defense capability and in the lives of all our citizens, America is a far better and more connected place because of work that originated from this office since its inception.

The information revolution, sparked by work supported by IPTO, created unprecedented changes in the way we work in this country and across the entire world. Studies cite dramatic increases in productivity. We can easily connect to virtually any information source in the world. Global communications is instant, and information is shared to a degree never before imagined.

With all this exciting history behind us, with our colleagues in MTO and other offices continuing to push the hardware research community to new heights, and with a robust microelectronics industry aggressively driving processor capabilities, why aren't we done? Isn't it time for the information processing side of the world to sit back and simply ride Moore's Law to bigger and better things?

Unfortunately, constant improvement in processor speed cuts two ways. While it gives us the opportunity to do bigger and better things, it inexorably leads us to do bigger and, well, bigger things. Everything we build on computational foundations keeps growing in complexity. The size of software systems is spiraling virtually out of control: 50-megabyte PC software applications are commonplace, and industrial applications easily soar above 5 million lines of code. And as we all know, this growth in complexity inevitably leads to new and very serious vulnerabilities.

From a security perspective, more complexity means more ways that intruders can find their way in. From a robustness point of view, more elements mean more ways that things can go wrong. From a maintenance point of view, more code means the cost of keeping things running just grows and grows. And from a training perspective, more people need to spend more time learning how to use computer-based systems.

These are problems of serious national importance. Because so much of our national infrastructure, both in defense and otherwise, rests on computer systems and software, we simply can't go on this way.

And I haven't even touched on something that affects all of us every day—the so-called "usability" of our systems. Despite marketing hype to the contrary, no computer system in widespread use today is really user-friendly. Some are, perhaps, user-tolerant, but unless the meaning of "friendly" has been warped beyond recognition recently, we still have a long way to go to get systems to adapt to people rather than the other way around.

All of this leads us to a clear conclusion: we need something dramatically different. We can't afford merely to increase the speed and capacity of our computer systems. We can't just do better software engineering. We need to change our perspective on computational systems and get off the current trajectory.

So, what do we propose as a solution here? Something that with a few moments' thought might seem fairly obvious to you. How many times have you watched something happen on your PC and just wished you could ask the stupid thing what it was doing and why it was doing it? How often have you wished that a system wouldn't simply remake the same mistake that it made the last time, but instead would learn by itself, or at least allow you to teach it how to avoid the problem? Despite our temptation to anthropomorphize and say that a system that is just sitting there doing nothing visible is “thinking”, we all know that it's not. Indeed, if today's systems were people, we would say that they are totally clueless. The sad part is that even if my PC had a clue, it wouldn't know what to do with it.

What we propose is nothing short of making our computer systems able to do what we all wish in these common circumstances: to reason, to learn, and to respond intelligently to things they've never encountered before. In other words, we want to transform them from systems that simply react to inputs into systems that are truly, in a word, cognitive.

Most formal and intuitive definitions tell us that cognition is about knowing. Our image of a cognitive system is one that can indeed know things and act on that knowledge. It can take explicit knowledge gleaned in a host of ways and go beyond it to important implicit knowledge through a variety of reasoning processes, ranging from pure and simple logical deduction—as in familiar syllogisms, like "Socrates is a man; all men are mortal; therefore, Socrates is mortal"—to what we might call plausible reasoning or back-of-the-envelope, educated guessing.

A truly cognitive system would be able to learn from its experience—as well as by being instructed—and perform better on day two than it did on day one. It would be able to explain what it was doing and why it was doing it. It would be reflective enough to know when it was heading down a blind alley or when it needed to ask for information that it simply couldn't get by further reasoning. And using these capabilities, a cognitive system would be robust in the face of surprises. It would be able to cope much more maturely with unanticipated circumstances than any current machine can. Such a system would finally be capable of being the cooperative, symbiotic partner that Licklider originally imagined.

This agenda is clearly ambitious and yet it is, perhaps, somewhat familiar. I can imagine that some of you are saying to yourselves, "Here we go again, those starry-eyed whiz kids at DARPA insist that the HAL-9000 is right around the corner when we know that artificial intelligence has never fulfilled its promise." But I submit to you four very important points:

First, AI and closely related disciplines, such as machine learning and speech processing, have made very significant strides in the last 20 years. There are industrial-strength versions of learning algorithms being applied right now to massive amounts of data. There are robust, speech recognition systems deployed in the public switched telephone network on a national scale. Multiple types of reasoning and their capabilities and limitations are understood in ways they never were before. AI techniques are integrated in many modern computer science subdisciplines, ranging from databases to operating systems to information retrieval.

Second, despite my somewhat tendentious remarks about Moore's Law a few minutes ago, the great strides that have been made in compressing more and more elements onto a chip are leading us to a point, fairly soon, where integrated circuits of the complexity of primate brains are actually foreseeable. I don't know when this will happen, but there is clearly tremendously greater computing power available to us than there was during the great wave of work on expert systems in the 1980s. This computing foundation is undeniably useful in getting us closer to human-level artificial intelligence.

Third, we are in the middle of an enormous revolution in neural science and understanding of the human brain. With new imaging and probing technology, the last decade has seemed to begin a golden age of insight into how the brain works. Progress on this front will be of enormous consequence to our ability to create artificial cognitive systems.

Finally, and more bluntly, we simply have no choice. We will fail if we keep doing more of the same. We have seen the costs and vulnerabilities of focusing only on speed and scale. We have struggled with networks that we can't configure by ourselves and interfaces that won't do what we want. It's time for a focused, serious effort in another direction.

And that is what IPTO intends to undertake.

So, how might we go about creating a new type of system that truly knows what it's doing? Let's start by reflecting for a moment on the kinds of capabilities that go into the human cognitive system and might provide a recipe for constructing an artificial one. Starting on the input side, we obviously need some sort of perceptual system. We are not talking simply about input transducers, but a system that does significant processing on the periphery before the input data is passed along to the core cognitive subsystem.

One of the remarkable things about natural perceptual systems is their ability to take what is an inordinate amount of raw sensor data, such as visual flow and constant, rich auditory and olfactory input, integrate and unify the key but disparate elements, and create from the result percepts that parcel the world into objects and discrete entities. This is all done, in a sense, without thinking. Yet the perception of objects, entities, and relationships seems very much influenced by a person's prior experience and what he or she knows. For example, I challenge you to try to stop interpreting your sensory input in discrete chunks. Look around you and just see continuous color or hear totally undistinguished sounds, without seeing people, walls and chandeliers or determining that this droning sound you hear is a person's voice. Maybe some of the Zen masters among you can do this, but I surely can't. And indeed, given the raw amount of data presented to us constantly, this is a wonderful and amazing thing.

Looking closely at this ability might give us some insights into how to notice important, low-frequency events-events that look roughly, but not exactly, like things we have seen before. Such capabilities are clearly important for national security.

On the flip side, we, of course, have effectors—muscles and activity—that are under our control. I won't say much about this now, but we have to take into account that even artificial cognitive systems will be embodied in some way and will, in fact, be what we might call "situated"; they are always operating in a context and, importantly, one that they can affect intentionally with actions.

Inside the core of a cognitive system, we see three types of processes operating most likely in parallel and with potentially many types of interaction. Let me quickly review these for you:

There are simple, relatively fast operations that we might call reactive. These are the things we do when we are operating, if you will, on autopilot, things we do without thinking. Simple reflexes, of course, fit this description, but there are many other things we do automatically. These can include what we've learned they don't just have to be innate and reflexive. For example, I don't think any of us were born with the knowledge to drive, but it is clear that much of the moment-to-moment activity of driving, once learned, is automatic and reactive.

At the true core of the cognitive system is a set of processes that we might call deliberative. These are processes that make up the bulk of what we mean by the word “thinking”. Trying to decide which direction to turn based on our current destination is a part of driving that is more deliberative. Trying to find the right way to express a thought in a sentence is deliberative. Planning a vacation or the day's errands or even a trip to the refrigerator during a commercial are deliberative acts. Deliberation definitely takes knowledge into account and, at least in some ways, knowledge derived in one type of deliberative process can be used in others.

There is one other type of process—we call this “reflective”—that we believe will be very important to the ultimate utility of cognitive systems. It's this type of process that most distinguishes higher animals from lower ones. For example, think of how at some point you might have discovered that you were getting nowhere in trying to solve a problem. You had the ability to stop the basic reasoning and reflect on alternative approaches. Or perhaps you consciously decided at some point that you were no longer interested in some mental activity and stopped it to move on to something else.

Self-awareness—or the ability to realize that we are individuals with certain capabilities, that we are here now, that we have past experiences different from those of others, and that we are conscious, reasoning entities—is an additional capability that makes reflection even more powerful.

The fact that reflective systems can stop what they are doing and, by stepping back from the situation, possibly get themselves out of a mental box is the reason we believe there is so much promise here. The higher-order cognitive processes of reflection and self-awareness could be key to creating systems that are not fragile in the presence of unforeseen inputs.

There is no doubt a lot more to say about the architecture of a cognitive system. For example, we fully expect there to be a long-term memory, which will contain what we think of as concepts and definitions as well as specific episodes and what would logically be called factual sentences.

Short-term memory, more of a buffer, might also be useful and might have different properties than long-term memory. And, of course, there are many types of connections between the layers of processes that we are contemplating, including ways for processes in one place to interrupt processes elsewhere, and for both deliberative and reflective processes to directly affect perception.But the details here need lots of work, and one of the true breakthroughs we'll be looking for is in the deep understanding of alternative architectures for cognition and how to match architectures ideally against problems.

Before we move ahead, I want to make one more observation. Notice that I did not postulate a specific learning component in the cognitive architecture.

Learning in the context of a full-blown cognitive system is not a unitary thing. There are many types of learning—whether they are based on some common mechanism, I can't tell you—but skill learning and language learning and discovering patterns in data and learning to build things are all different. And, of course, we all remember where we've been and put that kind of learning to good use in new situations. So our research in learning by machines will be varied, and the impact will be felt in different ways in different places, but in the end, very thoroughly throughout the architecture of the cognitive system.

In what I've said so far, I've focused solely on a single cognitive system. Many species, and especially humans, have been successful in this world because their members have not operated totally alone. They have the ability to form partnerships and teams and, as a result, accomplish goals that no individual acting alone conceivably could. This ranges from the simple, like lifting an object that no individual is capable of lifting alone, to the unbelievably complex, like successfully planning and executing an Apollo moon-landing mission or waging a protracted war against terrorism. As a result, one of the major elements of research that we plan to support is the creation and coordination of teams of cognitive systems.

Many of you are aware of the current yearly competitions in robot soccer. While we have come a long way from the first, uncoordinated efforts of teams of players unable to communicate with one another, there is still a long way to go. As we progress, we expect to even learn some things about teamwork and communication that can also be applied to human organizations. I'm sure many of us have been to meetings where, in a sense, the “collective IQ" of the meeting is somehow lower than the IQ of any of the individuals attending. Many of our implemented, multicomponent systems are like that now. We need to find a way to make a collective at least as smart, if not smarter, than its parts.

To recast what I've just intimated, IPTO will be focusing its attention on six core research areas over the next few years. First, we have

1. Computational perception
2. Representation and reasoning
3. Learning
4. Communication and interaction

These four capabilities make up the core cognitive part of a cognitive information processing system. In addition, we will be looking at the architecture of an individual cognitive system and how all these pieces can be integrated in the most effective way. Additionally, we clearly want to look at a fifth area:

5. Dynamic coordinated teams of cognitive systems

Finally, we care about the platforms on which our cognitive systems will be built. In that regard, our sixth area focuses on

6. Robust software and hardware infrastructure for building and maintaining cognitive systems and teams of cognitive systems.

This last topic demands a little attention of its own. While almost everything I have said so far has focused on the cognitive part of cognitive systems, it is important to realize that we are talking about systems. Among other things, robustness, by which I mean both fault-tolerance and security, is essential to these kinds of systems. IPTO's traditional constituency in areas such as networking will be vital to the success of our program. We also have great interest in the possibilities of new computing architectures coming out of the neuroscience world and from the bleeding edge of computing hardware design and fabrication.

While we have a great deal of work to do on the foundations of this new generation of computational systems, we certainly need to think of some of the specific application directions that will drive the focus of our work. It doesn't take much imagination to think of the myriad ways that "systems that know what they're doing" could make major differences in the operation of our military, the functioning of our Government, and the productivity of our daily lives.

Within our new office, we plan to focus on a small set of general application capabilities—we call them "application foundations"—that would support a very wide range of individual applications. Because we have so little time here, I will mention these only briefly, but over time you can expect to see much more detail from us.

The first of the application foundations is something we might call "adaptive networking."
I mean this both in the obvious sense and in a much broader sense.

In the simple interpretation, we are interested in networks whose components are smarter than they are now to allow for learning about evolving traffic patterns, error conditions, and emerging attacks. Ultimately, such networks should be able to ward off attacks by reasoning using deep knowledge about networking, communications, and prior forms of attack.

In the broader sense, we are interested in networks of entities that need to communicate and coordinate in some way. Continuing to build on the foundation of DARPA's DAML and CoABS work, we want to build practical frameworks for cognitive systems to introduce themselves, negotiate over shared goals, monitor team activity as it proceeds, and, ultimately, to succeed in creative ways to solve problems that as individuals they cannot.

A second basic capability that we need for important applications mirrors the remarks I made earlier about human perception: people have the ability to sort through extraordinarily large amounts of data in a reasonable period and to find the few bits and pieces that really matter. While this system is not flawless and does not necessarily work in one fell swoop, it is still quite remarkable.

Think of your ability to do a visual sweep of a room like this, guided simply by a high-level interest, such as "Are there any extraordinarily tall blond women here wearing large pointy hats?" I don't know if there are any, but it would take only a few seconds to get a pretty good idea if there are.

On top of our ability to find what amounts to tiny bits of information in a huge stream of data—needles in haystacks, if you will; or worse, needles in needle stacks—we have the ability to think through possible connections between these needles, and ultimately create a mental thread that might explain an ongoing event or even predict something we'd like to prevent in the near future.

This general ability to find the needles and then thread them, embodied in what we might call "perceptive agents," rests on many aspects of the operation of a cognitive system and is a very general capability that we need for a variety of information-intensive applications.

Another area of general concern is related to an aspect of cognitive systems that might be used to guide perception in an important way. We call this "strategic envisioning," and the idea is, in simple terms, to allow our cognitive systems to have an imagination.

Amidst the rhetoric following the events of last September, we heard the notion of failures of imagination, not being able to conceive what has never happened before. This takes some real, so-called out-of-the-box thinking. Given a cognitive system's ability to use knowledge and do hypothetical reasoning, it would be extremely useful for it to be able to do scenario planning and assess the likelihood of previously unimagined events. Many important applications need the ability to do hypothetical reasoning using extensive amounts of knowledge and would benefit from a flair for the imaginative.

A fourth application foundation is in the area of user interaction design and execution.

Self-aware entities—those that understand their own goals and the goals of other entities they are talking to—are able to adapt their output to suit their partners and the situation. To use a very mundane example, if I noticed that the people in the back of this room were squirming and then attended carefully to the sound of my own voice, I might detect that the microphone was not working. I could then adapt by speaking louder. Or I could double-check with the audience to see if I were audible. Further, if the audience was interested in having me illustrate a point in a certain way, it could simply ask, as is normal in a smaller, seminar-type setting.

What we need are user interfaces that are as responsive to the needs of their users as cooperative people can be. Users should be able to explain how and where they want to see something displayed and how something should sound.

This movement toward "form-fitting" interfaces (in other words, those that adapt to their users) is one that in and of itself could make a valuable contribution. In a way, this is the ultimate in user-centered design and a direction that we believe is very important. It is another key element of our program that mirrors Licklider's original dream.

Finally, it should not be necessary for every cognitive system built in the national interest to start with a tabula rasa. While it might not mirror ontogeny appropriately, a pre-installed knowledge base of important information of interest to a variety of systems would be a valuable asset.

This kind of "strategic knowledge bank" could cover a broad array of general concepts and facts about the operation of the Government and the military as well as factual descriptions of specific assets and their attributes. As a national corporate memory, this kind of knowledge base could be of value well beyond its use in seeding the memory of a cognitive system. Because so much important knowledge is not currently codified or easily available, a strategic knowledge reserve could be of vital importance in a time where other knowledge assets were unavailable or destroyed.

These five core application areas—networking and communication, information extraction, computational envisioning, form-fitting interfaces, and a national knowledge base—should form the foundation for a very wide set of applications.

The list of application foundations gives a pretty good idea of the type of practical application that we have our sights on. Over the course of the upcoming weeks and months, we will be formulating a plan that is likely to single out a small number of "challenge problems" that will help push the envelope of cognitive systems technology.

For the moment, you can imagine a continuum of applications ranging from relatively simple to extremely complex. On this spectrum, we might strive first to create software systems that were in some measure self-aware. This kind of system could help its creator in debugging and could extend its capability in a natural dialogue with its user. The kind of knowledge that such a seemingly simple system would need to possess is already a dramatic leap beyond current practice.

Next, we could imagine building on the basic adaptive networking capability I mentioned earlier and build a brand new, cognitive network. Such a network would be able to recognize interesting traffic patterns all on its own and adapt its routing in innovative ways. It could learn about novel forms of threat and attack over time and evolve its defenses to prevent damage. Given that it would have knowledge of routing, switching, connectivity, packet structures, and so on, it could also provide the basis for novel new communications capabilities that currently take extensive ad hoc network engineering.

Beyond the cognitive network, an application of great interest is an autonomous, perceiving agent. Such an independent program would be instructable in natural language so its user could explain how it wanted the agent to act, and the agent would figure out what to do and become what the user wanted. It could personalize its interaction with the user by inferring the user's preferences. And using perceptual capabilities of the sort we discussed earlier, it could find interesting threads and needles in the huge amounts of information flowing past it.

Finally, we want to create truly intelligent, multicomponent systems. These systems would potentially include multiple cognitive systems, humans, and noncognitive elements. By looking at the intelligence of the system as a whole, we could hope to produce large-scale systems that are substantially more effective and efficient than they are today. Redundancy of function and disconnects between role-players could be eliminated. And overall, the cost of the system could be dramatically reduced by its collective intelligence.

While our time together has really been too short to give you the kind of in-depth view of our new initiative that we would like, it is time for me to move off the stage so you can hear from some of our program managers about exciting work underway in our office. I want to mention that we plan to get off the mark extremely aggressively and, in particular, have already taken the step of issuing a Broad Agency Announcement on cognitive information processing technology. You can find a pointer to this BAA on the IPTO website, and we have hard copy versions of it here at DARPATech, which you can pick up at our exhibit booth.

We welcome the support of all of you in our efforts to have IPTO change the world yet again. If we are successful together, in another one or two dozen years, we will have finally realized much of Licklider's original dream of full-fledged, human-computer symbiosis, with computing truly accessible to all. Our computers will finally be able to be managed by everyday people. Systems will be easier to build and will last longer. We will move from the age of information to the age of cognition. Our systems will literally know what they're doing.

Thank you very much.